Reputations and Games

Sampath Kannan
Department of Computer and Information Science
University of Pennsylvania
November 4, 2009
ONR MURI Meeting

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
04 NOV 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Reputations and Games	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Pennsylvania,Computer and Information Science,Philadelphia,PA,19104**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

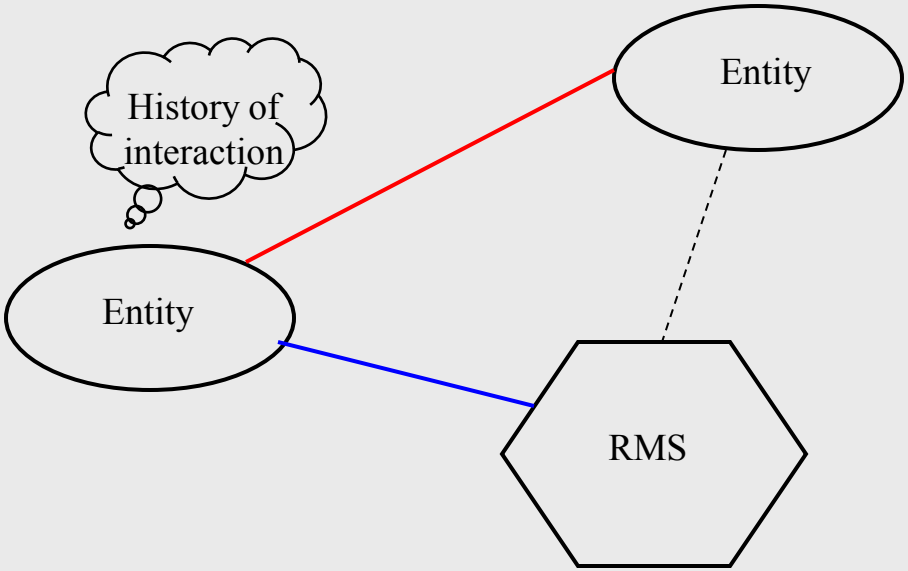
13. SUPPLEMENTARY NOTES
ONR MURI Review, Nov 2009.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**21**	

Reputation Management



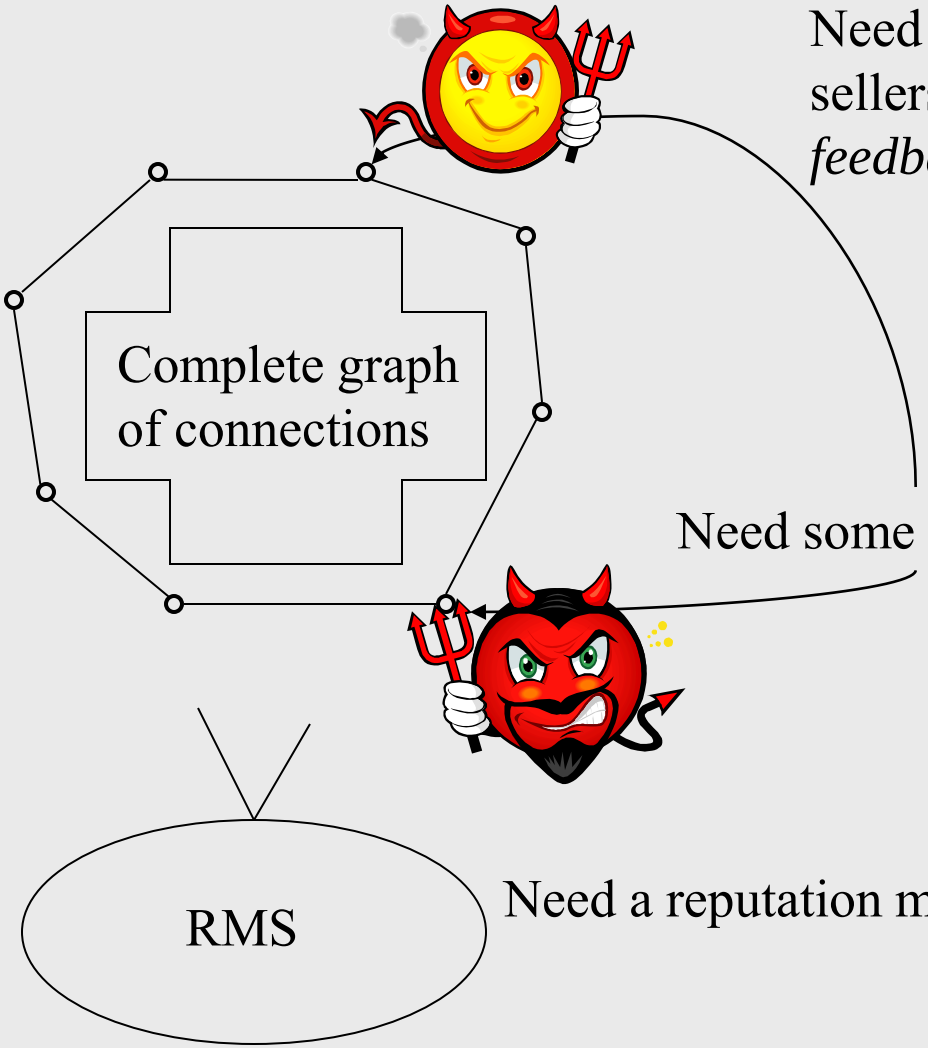
- Applicable to P2P
- "Degree" of Trust
- Evidence-based updates
- Amenable to Decision-Theoretic Framework

Although quantitative, these approaches are *ad hoc*, informal, and provide only simplistic capabilities.

Metrics – Challenge for security

- Metrics are hard for security, privacy, etc.
- Can we do better with reputation systems?
  - Would like measure of "value added" by RMS
  - Would allow us to compare RMS"s.
- We decided to try this for a simple framework

Simplest RMS framework?



Need a community – like buyers and sellers on eBay – with *transactions* and *feedback*

Need some bad apples

Need a reputation management system

What transactions?

- Simplest: “Abstract Client-Server Interactions”

- Client reports feedback to RMS.
- Bad guys could “frequently” provide bad service or bad feedback or both

Our transactions

- File sharing
  - More complex than “abstract transactions” because of side effect of file copy
  - Practically relevant
- Initially several copies of each file; probabilistically some are corrupted
- Good users
  - “clean up” i.e., remove bad files upon receipt with high probability
  - Provide honest feedback

Bad users

- Infinite variation possible in bad behavior
- To keep things simple we model bad users by two probabilities
  - Probability of clean-up α
  - Probability of honest feedback β
- What this doesn't allow:
  - Collusions
  - Targeted bad behavior

User Models

- A two-dimensional approach to behavior:
  - Cleanup (%): Upon reception of an invalid file, how likely is user to remove that file?
  - Honesty (%): With what probability will a user provide honest feedback

User Type	Cleanup	Honesty	Source
Good	90%-100%	100%	BEST
Purely Malicious	0%-10%	0%	WORST
Feedback Malicious	90%-100%	0%	RAND
Malicious Provider	0-10%	100%	WORST
Disguised Malicious	50%-100%	50%-100%	RAND
Sybil	Sybil users participate in 1 transaction then create a new „account".		WORST

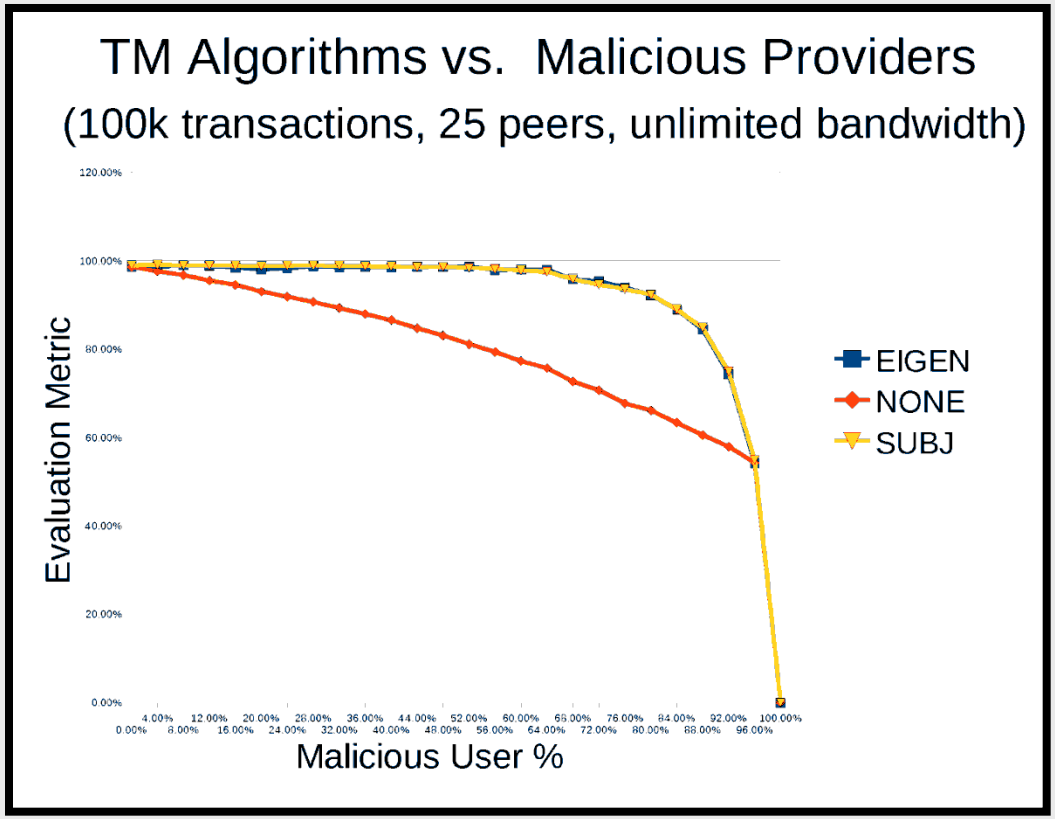
Note: "Source" dictates how Reputation values are used to choose a file-sender

TM systems analyzed

- None: The absence of TM, used for control runs
- EigenTrust by Hector Garcia-Molina et. al.
  - Globally convergent Reputation via matrix multiplication of normalized values
  - Convergence quick due to certain matrix properties
- Subjective Logic by Audun Jøsang et. al.
  - Triples of the form (belief, disbelief, uncertainty)
  - Transitive paths examined using „discount" and „consensus" logic operators
  - Reputation values correlate with beta-PDF functions

Metric & Results

- Metric: $$\frac{(\text{\# trans. with ``good'' recipients, resulting in trade of valid file})}{(\text{\# trans. attempted by ``good'' recepients})}$$

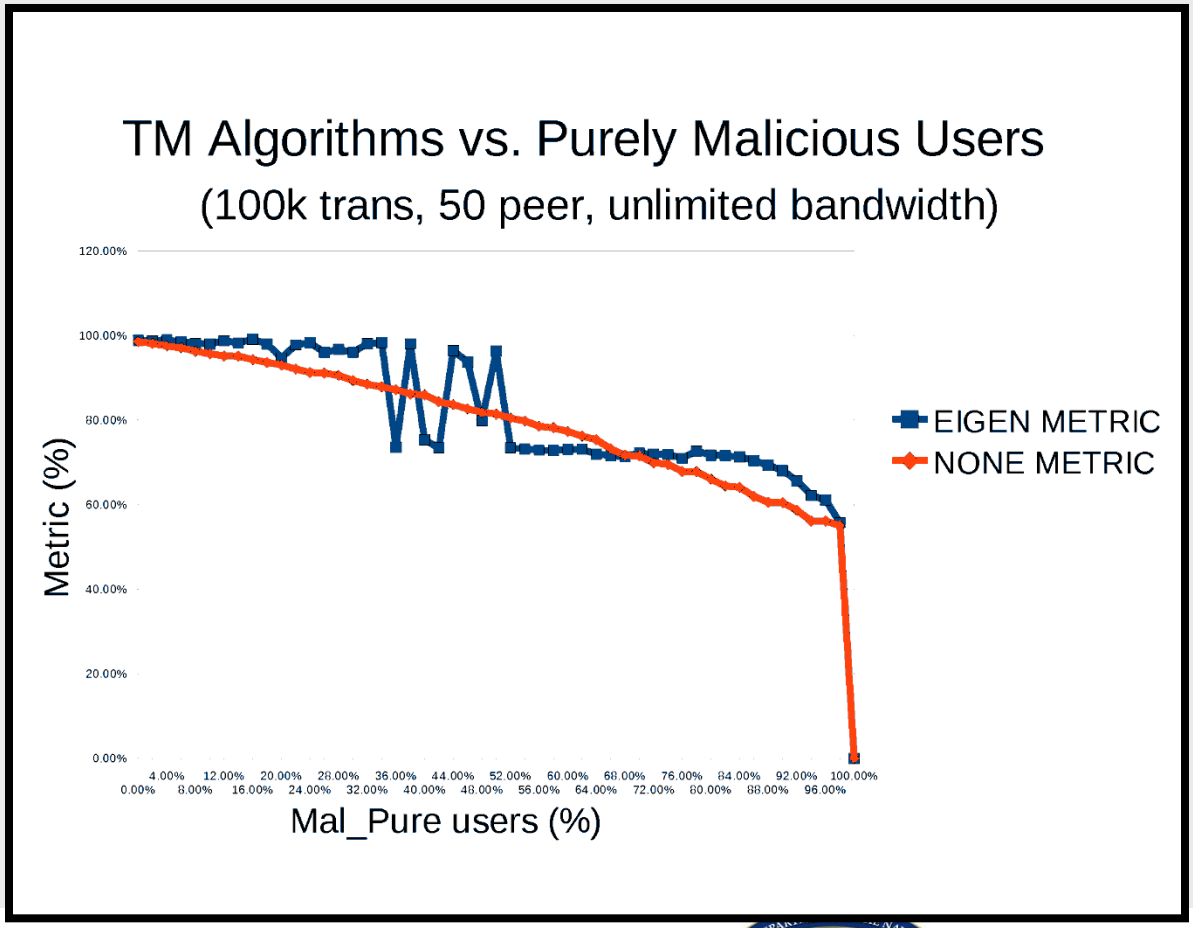


TM works well

A More Complex Example

More interesting is when users are bad, lie about their behavior, and have other bad peers lie on their behalf (as at right).

EigenTrust in particular demonstrates some very interesting properties under varying number of bad users (a topic currently under study).



Future Work (1)

- Adversary (BOTMaster) recruits nodes dynamically
  - Limited number of bad nodes
  - Dynamic trustworthiness - how incorruptible is a node?
- Collusions/Targeted attacks
  - Who do the bad nodes give bad feedback to? Bad files to?
- Correlations
  - Do corrupting nodes share bad files with corrupted nodes?

CS and Economics

Network of Agents

CS View

- Repeated interactions between sets of players
- Good players follow protocols
- Bad players adversarial: seek to inflict harm while escaping detection
- Goal: Prevent harm; maximize system utility

Econ View

- Repeated games with strategy-based payoffs
- All players "self-interested"
- What do "good" and "bad" mean?
- Goal: Maximize social welfare (at least welfare of "good" players)

Payoffs in Repeated Games (Economics)

Player *i* plays game at each time instant. At time *t* receives payoff $f_i(t)$.

Discounted payoff to player *i*: $(1-d)\ \Sigma\ d^t f_i(t)$

(Also possible to consider average payoff per game if limit exists… but discounted makes more sense usually.)

Repeated vs One-Shot Game: Example

- Prisoner"s Dilemma

	Confess	Silent
Confess	-6	-1
Silent	-9	-2

Payoffs to row player;
Symmetrically to column player

One-Shot Nash Equilibrium

Both Confess

Repeated Game Nash Equilibria

1) Grim strategy: Silent until opponent reneges
Confess thereafter
(Equilibrium if d high enough)
2) Both confess always

Good vs Bad in Econ View

- Games have multiple equilibria – bad players drive towards equilibria with low social welfare?
- Bad players have a small value of ; they don‟t care about future payoffs, and hence about “reputations”?
- alpha-altruism (New concept.) Player wants to optimize her own payoff + alpha(payoff to other players). (½-altruistic player payoff for (C,C) = -9)
- Bad Players may have evolutionary strategies to achieve objectives and mask behavior

Reputation Manager Outline

- Monitors interactions/games by one of:
  - Observing strategies played by all players
  - Observing payoffs to each player
  - Receiving feedback from players about strategies/payoffs
- Player strategies are function of advise from Reputation Manager and past history

Issues for Reputation Manager

- Thwart adversaries; enhance experience for good players
- Identify players who are bad in the senses defined above and “warn” good players
- Prevent following problems
  - Whitewash: acquire new identity after bad behavior
  - Phantom feedback: acquire multiple identities and provide spurious feedback to skew reputations
  - No feedback: fail to provide feedback when due

Evaluating Reputation Managers

- CS View:
  - Prevent harmful attacks
  - Make system available/useful to good players
- Econ View:
  - Increase total welfare to good players
  - Possibly enforce fair sharing of welfare among good players?

Future Work (2)

- Altruistic Players: An exciting new model? Explore
- Can reputation managers identify the altruism parameter of each player in an arbitrary game? Based on what observations?
- What if a player‟s degree of altruism is altered by nature of opponent?
- Analyze games under other notions of “badness”
- Reconcile Econ view to real systems?? Where do we get payoffs, lists of strategies from?